

PCTWORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C12N 15/12, C07K 14/47, 16/18, G01N 33/68, C12Q 1/68		A1	(11) International Publication Number: WO 00/05364 (43) International Publication Date: 3 February 2000 (03.02.00)		
(21) International Application Number: PCT/GB99/02360		(81) Designated States: JP, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).			
(22) International Filing Date: 21 July 1999 (21.07.99)		Published <i>With international search report.</i>			
(30) Priority Data: 9816024.5 22 July 1998 (22.07.98) GB 9916898.1 19 July 1999 (19.07.99) GB					
<p>(71) Applicant: SMITHKLINE BEECHAM PLC [GB/GB]; New Horizons Court, Brentford, Middlesex TW8 9EP (GB).</p> <p>(72) Inventors: MICHALOVICH, David; SmithKline Beecham Pharmaceuticals, New Frontiers Science Park South, Third Avenue, Harlow, Essex CM19 5AW (GB). PRINJHA, Rabinder, Kumar; SmithKline Beecham Pharmaceuticals, Third Avenue, Harlow, Essex CM19 5AW (GB).</p> <p>(74) Agent: CONNELL, Anthony, Christopher; SmithKline Beecham, Corporate Intellectual Property, Two New Horizons Court, Brentford, Middlesex TW8 9EP (GB).</p>					
(54) Title: PROTEIN SIMILAR TO NEUROENDOCRINE-SPECIFIC PROTEIN, AND ENCODING CDNA					
<p>(57) Abstract</p> <p>MAGI polypeptides and polynucleotides and methods for producing such polypeptides by recombinant techniques are disclosed. Also disclosed are methods for utilizing MAGI polypeptides and polynucleotides in diagnostic assays.</p>					

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon	KR	Republic of Korea	PL	Poland		
CN	China	KZ	Kazakhstan	PT	Portugal		
CU	Cuba	LC	Saint Lucia	RO	Romania		
CZ	Czech Republic	LI	Liechtenstein	RU	Russian Federation		
DE	Germany	LK	Sri Lanka	SD	Sudan		
DK	Denmark	LR	Liberia	SE	Sweden		
EE	Estonia			SG	Singapore		

PROTEIN SIMILAR TO NEUROENDOCRINE-SPECIFIC PROTEIN, AND ENCODING CDNA

Field of the Invention

This invention relates to newly identified polypeptides and polynucleotides encoding such polypeptides, to their use in diagnosis and in identifying compounds that may be agonists, antagonists that are potentially useful in therapy, and to production of such polypeptides and polynucleotides.

Background of the Invention

The drug discovery process is currently undergoing a fundamental revolution as it embraces "functional genomics", that is, high throughput genome- or gene-based biology. This approach as a means to identify genes and gene products as therapeutic targets is rapidly superceding earlier approaches based on "positional cloning". A phenotype, that is a biological function or genetic disease, would be identified and this would then be tracked back to the responsible gene, based on its genetic map position.

Functional genomics relies heavily on high-throughput DNA sequencing technologies and the various tools of bioinformatics to identify gene sequences of potential interest from the many molecular biology databases now available. There is a continuing need to identify and characterise further genes and their related polypeptides/proteins, as targets for drug discovery.

Summary of the Invention

The present invention relates to MAGI, in particular MAGI polypeptides and MAGI polynucleotides, recombinant materials and methods for their production. Such polypeptides and polynucleotides are of interest in relation to methods of treatment of certain diseases, including, but not limited to, neuropathies, spinal injury, neuronal degeneration, neuromuscular disorders, psychiatric disorders and developmental disorders, cancer, stroke and inflammatory disorders, hereinafter referred to as "diseases of the invention". In a further aspect, the invention relates to methods for identifying agonists and antagonists (e.g., inhibitors) using the materials provided by the invention, and treating conditions associated with MAGI imbalance with the identified compounds. In a still further aspect, the invention relates to diagnostic assays for detecting diseases associated with inappropriate MAGI activity or levels.

Description of the Invention

In a first aspect, the present invention relates to MAGI polypeptides. Such polypeptides include

- (a) an isolated polypeptide encoded by a polynucleotide comprising the sequence of SEQ ID NO:1 or SEQ ID NO:5;
- (b) an isolated polypeptide comprising a polypeptide sequence having at least 95%, 96%, 97%, 98% or 99% identity to the polypeptide sequence of SEQ ID NO:2 or SEQ ID NO:6;
- (c) an isolated polypeptide comprising the polypeptide sequence of SEQ ID NO:2 or SEQ ID NO:6;
- (d) an isolated polypeptide having at least 95%, 96%, 97%, 98%, or 99% identity to the polypeptide sequence of SEQ ID NO:2 or SEQ ID NO:6;
- (e) the polypeptide sequence of SEQ ID NO:2 or SEQ ID NO:6; and
- (f) an isolated polypeptide having or comprising a polypeptide sequence that has an Identity Index of 0.95, 0.96, 0.97, 0.98, or 0.99 compared to the polypeptide sequence of SEQ ID NO:2 or SEQ ID NO:6;
- (g) fragments and variants of such polypeptides in (a) to (f).

Polypeptides of the present invention are believed to be members of the reticulon family of polypeptides. They are therefore of interest because members of this family have been shown to display prominent, but not exclusive, expression in cells of the nervous system. Expression of one isoform of these polypeptides, neuroendocrine specific protein C (NSP-C) (J. Hens et al. Cell Tissue Res. 292:229-237, 1998), has been

5 shown to correlate with neuronal differentiation. Alternative splicing of the genes of this family of polypeptides is known to generate differentially expressed isoforms with overlapping and distinct functions in different tissues (Geisler et al. Mamm. Genome 9:164-173, 1998). Amino acid similarity between members of this family of polypeptides with fragments of a high-molecular weight protein purified from bovine spinal cord (Spillmann et al. J. Biol. Chem. 273:15487-15493, 1998) indicates a
10 potential role in axonal growth inhibition role for these proteins. Similarly, expression of neuroendocrine specific protein A (NSP-A) in specific cancerous cells (Senden et al. Histochem. Cell Biol. 108:155-165, 1997) may indicate a potential use of these polypeptides in the diagnosis and or treatment of cancers.

The detection of polynucleotides comprising portions of MAGI in human fetal brain and human adult spinal cord cDNA and an abundant >5kb mRNA isoform in human adult brain together with

15 different transcripts potentially arising by alternative splicing in heart, lung, liver, kidney and skeletal muscle suggests that MAGI isoforms may similarly serve overlapping and distinct functions in these tissues. By analogy with the semaphorin family of neurite-modulatory polypeptides it might be postulated that these different isoforms would function in each of these tissues to control local innervation by distinct neuronal populations. Aberrant expression of specific isoforms within, for example, skeletal
20 muscle would be predicted to alter motor and sensory neuron innervation in diseases such as amyotrophic lateral sclerosis (ALS).

25 The presence of characteristic signature polypeptides and highly hydrophobic regions in the polypeptide of the present invention suggests that its expression in regions of the nervous system and in tissues forming boundaries for growth may modulate growth and pathfinding both during development and following pathological or injurious processes.

Expression of the polypeptide of the invention, fragments thereof or alternatively spliced variants of the polypeptide on the surface of cells either naturally, as a secreted protein or following release by cellular damage may act on other cells either through specific receptors or pathologically through non-specific interactions to modulate cell attachment, spreading, migration or growth. This inhibition might be
30 expected to be either reversible or permanent possibly resulting in cell death. .

The biological properties of the MAGI are hereinafter referred to as "biological activity of MAGI" or "MAGI activity". Preferably, a polypeptide of the present invention exhibits at least one biological activity of MAGI.

35 Polypeptides of the present invention also includes variants of the aforementioned polypeptides, including all allelic forms and splice variants. Such polypeptides vary from the reference polypeptide by insertions, deletions, and substitutions that may be conservative or non-conservative, or any combination thereof. Particularly preferred variants are those in which several, for instance from 50 to 30, from 30 to 20, from 20 to 10, from 10 to 5, from 5 to 3, from 3 to 2, from 2 to 1 or 1 amino acids are inserted, substituted, or deleted, in any combination.

40 Preferred fragments of polypeptides of the present invention include an isolated polypeptide comprising an amino acid sequence having at least 30, 50 or 100 contiguous amino acids from the amino acid sequence of SEQ ID NO: 2 or SEQ ID NO:6, or an isolated polypeptide comprising an amino acid

sequence having at least 30, 50 or 100 contiguous amino acids truncated or deleted from the amino acid sequence of SEQ ID NO: 2 or SEQ ID NO:6. Preferred fragments are biologically active fragments that mediate the biological activity of MAGI, including those with a similar activity or an improved activity, or with a decreased undesirable activity. Also preferred are those fragments that are antigenic or immunogenic in an animal, especially in a human.

5 Fragments of the polypeptides of the invention may be employed for producing the corresponding full-length polypeptide by peptide synthesis; therefore, these variants may be employed as intermediates for producing the full-length polypeptides of the invention. The polypeptides of the present invention may be in the form of the "mature" protein or may be a part of a larger protein such as a precursor or a fusion 10 protein. It is often advantageous to include an additional amino acid sequence that contains secretory or leader sequences, pro-sequences, sequences that aid in purification, for instance multiple histidine residues, or an additional sequence for stability during recombinant production.

15 Polypeptides of the present invention can be prepared in any suitable manner, for instance by isolation from naturally occurring sources, from genetically engineered host cells comprising expression systems (*vide infra*) or by chemical synthesis, using for instance automated peptide synthesisers, or a combination of such methods.. Means for preparing such polypeptides are well understood in the art.

In a further aspect, the present invention relates to MAGI polynucleotides. Such polynucleotides include:

- 20 (a) an isolated polynucleotide comprising a polynucleotide sequence having at least 95%, 96%, 97%, 98%, or 99% identity to the polynucleotide sequence of SEQ ID NO:1 or SEQ ID NO:5;
- (b) an isolated polynucleotide comprising the polynucleotide of SEQ ID NO:1 or SEQ ID NO:5;
- (c) an isolated polynucleotide having at least 95%, 96%, 97%, 98%, or 99% identity to the polynucleotide of SEQ ID NO:1 or SEQ ID NO:5;
- 25 (d) the isolated polynucleotide of SEQ ID NO:1 or SEQ ID NO:5;
- (e) an isolated polynucleotide comprising a polynucleotide sequence encoding a polypeptide sequence having at least 95%, 96%, 97%, 98%, or 99% identity to the polypeptide sequence of SEQ ID NO:2 or SEQ ID NO:6;
- (f) an isolated polynucleotide comprising a polynucleotide sequence encoding the polypeptide of SEQ ID NO:2 or SEQ ID NO:6;
- 30 (g) an isolated polynucleotide having a polynucleotide sequence encoding a polypeptide sequence having at least 95%, 96%, 97%, 98%, or 99% identity to the polypeptide sequence of SEQ ID NO:2 or SEQ ID NO:6;
- (h) an isolated polynucleotide encoding the polypeptide of SEQ ID NO:2 or SEQ ID NO:6;
- 35 (i) an isolated polynucleotide having or comprising a polynucleotide sequence that has an Identity Index of 0.95, 0.96, 0.97, 0.98, or 0.99 compared to the polynucleotide sequence of SEQ ID NO:1 or SEQ ID NO:5;
- (j) an isolated polynucleotide having or comprising a polynucleotide sequence encoding a polypeptide sequence that has an Identity Index of 0.95, 0.96, 0.97, 0.98, or 0.99 compared to the polypeptide sequence of SEQ ID NO:2 or SEQ ID NO:6; and
- 40 polynucleotides that are fragments and variants of the above mentioned polynucleotides or that are complementary to above mentioned polynucleotides, over the entire length thereof.

Preferred fragments of polynucleotides of the present invention include an isolated polynucleotide comprising an nucleotide sequence having at least 15, 30, 50 or 100 contiguous nucleotides from the sequence of SEQ ID NO: 1 or SEQ ID NO:5, or an isolated polynucleotide comprising an sequence having at least 30, 50 or 100 contiguous nucleotides truncated or deleted from the sequence of SEQ ID NO: 1 or SEQ ID NO:5.

Preferred variants of polynucleotides of the present invention include splice variants, allelic variants, and polymorphisms, including polynucleotides having one or more single nucleotide polymorphisms (SNPs).

Polynucleotides of the present invention also include polynucleotides encoding polypeptide variants that comprise the amino acid sequence of SEQ ID NO:2 or SEQ ID NO:6 and in which several, for instance from 50 to 30, from 30 to 20, from 20 to 10, from 10 to 5, from 5 to 3, from 3 to 2, from 2 to 1 or 1 amino acid residues are substituted, deleted or added, in any combination.

In a further aspect, the present invention provides polynucleotides that are RNA transcripts of the DNA sequences of the present invention. Accordingly, there is provided an RNA polynucleotide that:

- 15 (a) comprises an RNA transcript of the DNA sequence encoding the polypeptide of SEQ ID NO:2 or SEQ ID NO:6;
- (b) is the RNA transcript of the DNA sequence encoding the polypeptide of SEQ ID NO:2 or SEQ ID NO:6;
- (c) comprises an RNA transcript of the DNA sequence of SEQ ID NO:1 or SEQ ID NO:5; or
- 20 (d) is the RNA transcript of the DNA sequence of SEQ ID NO:1 or SEQ ID NO:5; and RNA polynucleotides that are complementary thereto.

The polynucleotide sequence of SEQ ID NO:1 shows homology with Human neuroendocrine-specific protein C (Roebroek et al., J. Biol. Chem. 268: 13439-13447,1993). The polynucleotide sequence of SEQ ID NO:1 is a cDNA sequence that encodes the polypeptide of SEQ ID NO:2. The polynucleotide sequence encoding the polypeptide of SEQ ID NO:2 may be identical to the polypeptide encoding sequence of SEQ ID NO:1 or it may be a sequence other than SEQ ID NO:1, which, as a result of the redundancy (degeneracy) of the genetic code, also encodes the polypeptide of SEQ ID NO:2. The polypeptide of the SEQ ID NO:2 is related to other proteins of the reticulon family, having homology and/or structural similarity with Human neuroendocrine-specific protein C (Roebroek et al., J. Biol. Chem. 268: 13439-13447,1993). The polynucleotide of SEQ ID NO:5 and the polypeptide encoded thereby, SEQ ID NO:6, is a variant of the polynucleotide of SEQ ID NO:1 and it's encoded polypeptide, SEQ ID NO:2.

Preferred polypeptides and polynucleotides of the present invention are expected to have, *inter alia*, similar biological functions/properties to their homologous polypeptides and polynucleotides. Furthermore, preferred polypeptides and polynucleotides of the present invention have at least one MAGI activity.

The present invention also relates to partial or other polynucleotide and polypeptide sequences which were first identified prior to the determination of the corresponding full length sequences of SEQ ID NO:1, SEQ ID NO:5, SEQ ID NO:2 and SEQ ID NO:6.

Accordingly, in a further aspect, the present invention provides for an isolated polynucleotide which:

- (a) comprises a nucleotide sequence which has at least 95% identity, preferably at least 97-99% identity to SEQ ID NO:3 over the entire length of SEQ ID NO:3;

- (b) has a nucleotide sequence which has at least 95% identity, preferably at least 97-99% identity, to SEQ ID NO:3 over the entire length of SEQ ID NO:3;
- (c) comprises the polynucleotide of SEQ ID NO:3; or
- (d) has a nucleotide sequence encoding a polypeptide which has at least 95% identity, even more preferably at least 97-99% identity, to the amino acid sequence of SEQ ID NO:4, over the entire length of SEQ ID NO:4;
- 5 as well as the polynucleotide of SEQ ID NO:3.

The present invention further provides for a polypeptide which:

- (a) comprises an amino acid sequence which has at least 95% identity, preferably at least 97-99% identity, to that of SEQ ID NO:4 over the entire length of SEQ ID NO:4;
- 10 (b) has an amino acid sequence which is at least 95% identity, preferably at least 97-99% identity, to the amino acid sequence of SEQ ID NO:4 over the entire length of SEQ ID NO:4;
- (c) comprises the amino acid of SEQ ID NO:4; and
- (d) is the polypeptide of SEQ ID NO:4;
- 15 as well as polypeptides encoded by a polynucleotide comprising the sequence contained in SEQ ID NO:3.

The nucleotide sequence of SEQ ID NO:3 and the peptide sequence encoded thereby are derived from EST (Expressed Sequence Tag) sequences. It is recognised by those skilled in the art that there will inevitably be some nucleotide sequence reading errors in EST sequences (see Adams, M.D. *et al*, Nature 377 (supp) 3, 1995). Accordingly, the nucleotide sequence of SEQ ID NO:3 and the peptide sequence encoded therefrom are therefore subject to the same inherent limitations in sequence accuracy.

20 Furthermore, the peptide sequence encoded by SEQ ID NO:3 comprises a region of identity or close homology and/or close structural similarity (for example a conservative amino acid difference) with the closest homologous or structurally similar protein.

Polynucleotides of the present invention may be obtained using standard cloning and screening techniques from a cDNA library derived from mRNA in cells of human fetal brain and spinal cord, (see for instance, Sambrook *et al.*, Molecular Cloning: A Laboratory Manual. 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989)). Polynucleotides of the invention can also be obtained from natural sources such as genomic DNA libraries or can be synthesized using well known and commercially available techniques.

When polynucleotides of the present invention are used for the recombinant production of polypeptides of the present invention, the polynucleotide may include the coding sequence for the mature polypeptide, by itself, or the coding sequence for the mature polypeptide in reading frame with other coding sequences, such as those encoding a leader or secretory sequence, a pre-, or pro- or prepro- protein sequence, or other fusion peptide portions. For example, a marker sequence that facilitates purification of the fused polypeptide can be encoded. In certain preferred embodiments of this aspect of the invention, the marker sequence is a hexa-histidine peptide, as provided in the pQE vector (Qiagen, Inc.) and described in Gentz *et al.*, Proc Natl Acad Sci USA (1989) 86:821-824, or is an HA tag. The polynucleotide may also contain non-coding 5' and 3' sequences, such as transcribed, non-translated sequences, splicing and polyadenylation signals, ribosome binding sites and sequences that stabilize mRNA.

40 Polynucleotides that are identical, or have sufficient identity to a polynucleotide sequence of SEQ ID NO:1 or SEQ ID NO:5, may be used as hybridization probes for cDNA and genomic DNA or as primers for a nucleic acid amplification reaction (for instance, PCR). Such probes and primers may be used to isolate

full-length cDNAs and genomic clones encoding polypeptides of the present invention and to isolate cDNA and genomic clones of other genes (including genes encoding paralogs from human sources and orthologs and paralogs from species other than human) that have a high sequence similarity to SEQ ID NO:1 or SEQ ID NO:5, typically at least 95% identity. Preferred probes and primers will generally comprise at least 15 nucleotides, preferably, at least 30 nucleotides and may have at least 50, if not at least 100 nucleotides. Particularly preferred probes will have between 30 and 50 nucleotides. Particularly preferred primers will have between 20 and 25 nucleotides.

A polynucleotide encoding a polypeptide of the present invention, including homologs from species other than human, may be obtained by a process comprising the steps of screening a library under stringent hybridization conditions with a labeled probe having the sequence of SEQ ID NO: 1 or SEQ ID NO:5 or a fragment thereof, preferably of at least 15 nucleotides; and isolating full-length cDNA and genomic clones containing said polynucleotide sequence. Such hybridization techniques are well known to the skilled artisan. Preferred stringent hybridization conditions include overnight incubation at 42°C in a solution comprising: 50% formamide, 5xSSC (150mM NaCl, 15mM trisodium citrate), 50 mM sodium phosphate (pH7.6), 5x Denhardt's solution, 10 % dextran sulfate, and 20 microgram/ml denatured, sheared salmon sperm DNA; followed by washing the filters in 0.1x SSC at about 65°C. Thus the present invention also includes isolated polynucleotides, preferably with a nucleotide sequence of at least 100, obtained by screening a library under stringent hybridization conditions with a labeled probe having the sequence of SEQ ID NO:1 or SEQ ID NO:5 or a fragment thereof, preferably of at least 15 nucleotides.

The skilled artisan will appreciate that, in many cases, an isolated cDNA sequence will be incomplete, in that the region coding for the polypeptide does not extend all the way through to the 5' terminus. This is a consequence of reverse transcriptase, an enzyme with inherently low "processivity" (a measure of the ability of the enzyme to remain attached to the template during the polymerisation reaction), failing to complete a DNA copy of the mRNA template during first strand cDNA synthesis.

There are several methods available and well known to those skilled in the art to obtain full-length cDNAs, or extend short cDNAs, for example those based on the method of Rapid Amplification of cDNA ends (RACE) (see, for example, Frohman et al., Proc Nat Acad Sci USA 85, 8998-9002, 1988). Recent modifications of the technique, exemplified by the Marathon (trade mark) technology (Clontech Laboratories Inc.) for example, have significantly simplified the search for longer cDNAs. In the Marathon (trade mark) technology, cDNAs have been prepared from mRNA extracted from a chosen tissue and an 'adaptor' sequence ligated onto each end. Nucleic acid amplification (PCR) is then carried out to amplify the "missing" 5' end of the cDNA using a combination of gene specific and adaptor specific oligonucleotide primers. The PCR reaction is then repeated using 'nested' primers, that is, primers designed to anneal within the amplified product (typically an adaptor specific primer that anneals further 3' in the adaptor sequence and a gene specific primer that anneals further 5' in the known gene sequence). The products of this reaction can then be analysed by DNA sequencing and a full-length cDNA constructed either by joining the product directly to the existing cDNA to give a complete sequence, or carrying out a separate full-length PCR using the new sequence information for the design of the 5' primer.

Recombinant polypeptides of the present invention may be prepared by processes well known in the art from genetically engineered host cells comprising expression systems. Accordingly, in a further aspect,

the present invention relates to expression systems comprising a polynucleotide or polynucleotides of the present invention, to host cells which are genetically engineered with such expression systems and to the production of polypeptides of the invention by recombinant techniques. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention.

For recombinant production, host cells can be genetically engineered to incorporate expression systems or portions thereof for polynucleotides of the present invention. Polynucleotides may be introduced into host cells by methods described in many standard laboratory manuals, such as Davis et al. *Basic Methods in Molecular Biology* (1986) and Sambrook *et al.* (*ibid*). Preferred methods of introducing polynucleotides into host cells include, for instance, calcium phosphate transfection, DEAE-dextran mediated transfection, transvection, microinjection, cationic lipid-mediated transfection, electroporation, transduction, scrape loading, ballistic introduction or infection.

Representative examples of appropriate hosts include bacterial cells, such as *Streptococci*, *Staphylococci*, *E. coli*, *Streptomyces* and *Bacillus subtilis* cells; fungal cells, such as yeast cells and *Aspergillus* cells; insect cells such as *Drosophila S2* and *Spodoptera Sf9* cells; animal cells such as CHO, COS, HeLa, C127, 3T3, BHK, HEK 293 and Bowes melanoma cells; and plant cells.

A great variety of expression systems can be used, for instance, chromosomal, episomal and virus-derived systems, e.g., vectors derived from bacterial plasmids, from bacteriophage, from transposons, from yeast episomes, from insertion elements, from yeast chromosomal elements, from viruses such as baculoviruses, papova viruses, such as SV40, vaccinia viruses, adenoviruses, fowl pox viruses, pseudorabies viruses and retroviruses, and vectors derived from combinations thereof, such as those derived from plasmid and bacteriophage genetic elements, such as cosmids and phagemids. The expression systems may contain control regions that regulate as well as engender expression. Generally, any system or vector that is able to maintain, propagate or express a polynucleotide to produce a polypeptide in a host may be used. The appropriate polynucleotide sequence may be inserted into an expression system by any of a variety of well-known and routine techniques, such as, for example, those set forth in Sambrook *et al.*, (*ibid*). Appropriate secretion signals may be incorporated into the desired polypeptide to allow secretion of the translated protein into the lumen of the endoplasmic reticulum, the periplasmic space or the extracellular environment. These signals may be endogenous to the polypeptide or they may be heterologous signals.

If a polypeptide of the present invention is to be expressed for use in screening assays, it is generally preferred that the polypeptide be produced at the surface of the cell. In this event, the cells may be harvested prior to use in the screening assay. If the polypeptide is secreted into the medium, the medium can be recovered in order to recover and purify the polypeptide. If produced intracellularly, the cells must first be lysed before the polypeptide is recovered.

Polypeptides of the present invention can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Most preferably, high performance liquid chromatography is employed for purification. Well known techniques for refolding proteins may be employed to regenerate active conformation when the polypeptide is denatured during intracellular synthesis, isolation and/or purification.

Polynucleotides of the present invention may be used as diagnostic reagents, through detecting mutations in the associated gene. Detection of a mutated form of the gene characterised by the polynucleotide of SEQ ID NO:1 or SEQ ID NO:5 in the cDNA or genomic sequence and which is associated with a dysfunction will provide a diagnostic tool that can add to, or define, a diagnosis of a disease, or 5 susceptibility to a disease, which results from under-expression, over-expression or altered spatial or temporal expression of the gene. Individuals carrying mutations in the gene may be detected at the DNA level by a variety of techniques well known in the art.

Nucleic acids for diagnosis may be obtained from a subject's cells, such as from blood, urine, saliva, tissue biopsy or autopsy material. The genomic DNA may be used directly for detection or it may be 10 amplified enzymatically by using PCR, preferably RT-PCR, or other amplification techniques prior to analysis. RNA or cDNA may also be used in similar fashion. Deletions and insertions can be detected by a change in size of the amplified product in comparison to the normal genotype. Point mutations can be identified by hybridizing amplified DNA to labeled MAGI nucleotide sequences. Perfectly matched 15 sequences can be distinguished from mismatched duplexes by RNase digestion or by differences in melting temperatures. DNA sequence difference may also be detected by alterations in the electrophoretic mobility of DNA fragments in gels, with or without denaturing agents, or by direct DNA sequencing (see, for instance, Myers *et al.*, *Science* (1985) 230:1242). Sequence changes at specific locations may also be revealed by nuclease protection assays, such as RNase and S1 protection or the chemical cleavage method (see Cotton *et 20* al., *Proc Natl Acad Sci USA* (1985) 85: 4397-4401).

An array of oligonucleotides probes comprising MAGI polynucleotide sequence or fragments thereof can be constructed to conduct efficient screening of e.g., genetic mutations. Such arrays are 25 preferably high density arrays or grids. Array technology methods are well known and have general applicability and can be used to address a variety of questions in molecular genetics including gene expression, genetic linkage, and genetic variability, see, for example, M.Chee *et al.*, *Science*, 274, 610-613 (1996) and other references cited therein.

Detection of abnormally decreased or increased levels of polypeptide or mRNA expression may also be used for diagnosing or determining susceptibility of a subject to a disease of the invention. Decreased or increased expression can be measured at the RNA level using any of the methods well known in the art 30 for the quantitation of polynucleotides, such as, for example, nucleic acid amplification, for instance PCR, RT-PCR, RNase protection, Northern blotting and other hybridization methods. Assay techniques that can be used to determine levels of a protein, such as a polypeptide of the present invention, in a sample derived from a host are well-known to those of skill in the art. Such assay methods include radioimmunoassays, competitive-binding assays, Western Blot analysis and ELISA assays.

Thus in another aspect, the present invention relates to a diagnostic kit comprising:

35 (a) a polynucleotide of the present invention, preferably the nucleotide sequence of SEQ ID NO: 1 or SEQ ID NO:5, or a fragment or an RNA transcript thereof;

(b) a nucleotide sequence complementary to that of (a);

(c) a polypeptide of the present invention, preferably the polypeptide of SEQ ID NO:2 or SEQ ID NO:6 or a fragment thereof; or

40 (d) an antibody to a polypeptide of the present invention, preferably to the polypeptide of SEQ ID NO:2 or SEQ ID NO:6.

It will be appreciated that in any such kit, (a), (b), (c) or (d) may comprise a substantial component. Such a kit will be of use in diagnosing a disease or susceptibility to a disease, particularly diseases of the invention, amongst others.

5 The polynucleotide sequences of the present invention are valuable for chromosome localisation studies. The sequence is specifically targeted to, and can hybridize with, a particular location on an individual human chromosome. The mapping of relevant sequences to chromosomes according to the present invention is an important first step in correlating those sequences with gene associated disease. Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the
10 chromosome can be correlated with genetic map data. Such data are found in, for example, V. McKusick, Mendelian Inheritance in Man (available on-line through Johns Hopkins University Welch Medical Library). The relationship between genes and diseases that have been mapped to the same chromosomal region are then identified through linkage analysis (co-inheritance of physically adjacent genes). Precise human chromosomal localisations for a genomic sequence (gene fragment etc.) can be determined using
15 Radiation Hybrid (RH) Mapping (Walter, M. Spillett, D., Thomas, P., Weissenbach, J., and Goodfellow, P., (1994) A method for constructing radiation hybrid maps of whole genomes. *Nature Genetics* 7, 22-28). A number of RH panels are available from Research Genetics (Huntsville, AL, USA) e.g. the GeneBridge4 RH panel (*Hum Mol Genet* 1996 Mar;5(3):339-46) A radiation hybrid map of the human genome. Gyapay G, Schmitt K, Fizames C, Jones H, Vega-Czarny N, Spillett D, Muselet D, Prud'Homme
20 JF, Dib C, Auffray C, Morissette J, Weissenbach J, Goodfellow PN). To determine the chromosomal location of a gene using this panel, 93 PCRs are performed using primers designed from the gene of interest on RH DNAs. Each of these DNAs contains random human genomic fragments maintained in a hamster background (human / hamster hybrid cell lines). These PCRs result in 93 scores indicating the presence or absence of the PCR product of the gene of interest. These scores are compared with scores
25 created using PCR products from genomic sequences of known location. This comparison is conducted at <http://www.genome.wi.mit.edu/>. The gene of the present invention maps to human chromosome 2p21.

30 The polynucleotide sequences of the present invention are also valuable tools for tissue expression studies. Such studies allow the determination of expression patterns of polynucleotides of the present invention which may give an indication as to the expression patterns of the encoded polypeptides in tissues, by detecting the mRNAs that encode them. The techniques used are well known in the art and include *in situ* hybridisation techniques to clones arrayed on a grid, such as cDNA microarray hybridisation (Schena *et al.*, *Science*, 270, 467-470, 1995 and Shalon *et al.*, *Genome Res.*, 6, 639-645, 1996) and nucleotide amplification techniques such as PCR. A preferred method uses the TAQMAN (Trade mark) technology available from Perkin Elmer. Results from these studies can provide an indication of the normal function of the polypeptide in the organism. In addition, comparative studies of the normal expression pattern of mRNAs with that of mRNAs encoded by an alternative form of the same gene (for example, one having an alteration in polypeptide coding potential or a regulatory mutation) can provide valuable insights into the role of the polypeptides of the present invention, or that of inappropriate expression thereof in disease. Such
35 inappropriate expression may be of a temporal, spatial or simply quantitative nature. The polypeptides of the present invention are expressed in brain, heart, liver, skeletal muscle, pancreas and kidney, based on Northern blot data provided in figure 1.
40

A further aspect of the present invention relates to antibodies. The polypeptides of the invention or their fragments, or cells expressing them, can be used as immunogens to produce antibodies that are immunospecific for polypeptides of the present invention. The term "immunospecific" means that the 5 antibodies have substantially greater affinity for the polypeptides of the invention than their affinity for other related polypeptides in the prior art.

Antibodies generated against polypeptides of the present invention may be obtained by administering the polypeptides or epitope-bearing fragments, or cells to an animal, preferably a non-human animal, using routine protocols. For preparation of monoclonal antibodies, any technique which provides antibodies 10 produced by continuous cell line cultures can be used. Examples include the hybridoma technique (Kohler, G. and Milstein, C., *Nature* (1975) 256:495-497), the trioma technique, the human B-cell hybridoma technique (Kozbor *et al.*, *Immunology Today* (1983) 4:72) and the EBV-hybridoma technique (Cole *et al.*, *Monoclonal Antibodies and Cancer Therapy*, 77-96, Alan R. Liss, Inc., 1985).

Techniques for the production of single chain antibodies, such as those described in U.S. Patent No. 15 4,946,778, can also be adapted to produce single chain antibodies to polypeptides of this invention. Also, transgenic mice, or other organisms, including other mammals, may be used to express humanized antibodies.

The above-described antibodies may be employed to isolate or to identify clones expressing the 20 polypeptide or to purify the polypeptides by affinity chromatography. Antibodies against polypeptides of the present invention may also be employed to treat diseases of the invention, amongst others.

Polypeptides and polynucleotides of the present invention may also be used as vaccines. Accordingly, in a further aspect, the present invention relates to a method for inducing an immunological response in a mammal that comprises inoculating the mammal with a polypeptide of the present 25 invention, adequate to produce antibody and/or T cell immune response, including, for example, cytokine-producing T cells or cytotoxic T cells, to protect said animal from disease, whether that disease is already established within the individual or not. An immunological response in a mammal may also be induced by a method comprising delivering a polypeptide of the present invention *via* a vector directing expression of the polynucleotide and coding for the polypeptide *in vivo* in order to induce such an 30 immunological response to produce antibody to protect said animal from diseases of the invention. One way of administering the vector is by accelerating it into the desired cells as a coating on particles or otherwise. Such nucleic acid vector may comprise DNA, RNA, a modified nucleic acid, or a DNA/RNA hybrid. For use as a vaccine, a polypeptide or a nucleic acid vector will be normally provided as a vaccine 35 formulation (composition). The formulation may further comprise a suitable carrier. Since a polypeptide may be broken down in the stomach, it is preferably administered parenterally (for instance, subcutaneous, intramuscular, intravenous, or intradermal injection). Formulations suitable for parenteral administration include aqueous and non-aqueous sterile injection solutions that may contain anti-oxidants, buffers, bacteriostats and solutes that render the formulation isotonic with the blood of the recipient; and aqueous and non-aqueous sterile suspensions that may include suspending agents or thickening agents. 40 The formulations may be presented in unit-dose or multi-dose containers, for example, sealed ampoules and vials and may be stored in a freeze-dried condition requiring only the addition of the sterile liquid carrier immediately prior to use. The vaccine formulation may also include adjuvant systems for

enhancing the immunogenicity of the formulation, such as oil-in water systems and other systems known in the art. The dosage will depend on the specific activity of the vaccine and can be readily determined by routine experimentation.

5 Polypeptides of the present invention have one or more biological functions that are of relevance in one or more disease states, in particular the diseases of the invention hereinbefore mentioned. It is therefore useful to identify compounds that stimulate or inhibit the function or level of the polypeptide. Accordingly, in a further aspect, the present invention provides for a method of screening compounds to identify those that stimulate or inhibit the function or level of the polypeptide. Such methods identify
10 agonists or antagonists that may be employed for therapeutic and prophylactic purposes for such diseases of the invention as hereinbefore mentioned. Compounds may be identified from a variety of sources, for example, cells, cell-free preparations, chemical libraries, collections of chemical compounds, and natural product mixtures. Such agonists or antagonists so-identified may be natural or modified substrates, ligands,
15 receptors, enzymes, etc., as the case may be, of the polypeptide; a structural or functional mimetic thereof (see Coligan *et al.*, Current Protocols in Immunology 1(2):Chapter 5 (1991)) or a small molecule.

The screening method may simply measure the binding of a candidate compound to the polypeptide, or to cells or membranes bearing the polypeptide, or a fusion protein thereof, by means of a label directly or indirectly associated with the candidate compound. Alternatively, the screening method may involve measuring or detecting (qualitatively or quantitatively) the competitive binding of a
20 candidate compound to the polypeptide against a labeled competitor (e.g. agonist or antagonist). Further, these screening methods may test whether the candidate compound results in a signal generated by activation or inhibition of the polypeptide, using detection systems appropriate to the cells bearing the polypeptide. Inhibitors of activation are generally assayed in the presence of a known agonist and the effect on activation by the agonist by the presence of the candidate compound is observed. Further, the
25 screening methods may simply comprise the steps of mixing a candidate compound with a solution containing a polypeptide of the present invention, to form a mixture, measuring a MAGI activity in the mixture, and comparing the MAGI activity of the mixture to a control mixture which contains no candidate compound.

Polypeptides of the present invention may be employed in conventional low capacity screening
30 methods and also in high-throughput screening (HTS) formats. Such HTS formats include not only the well-established use of 96- and, more recently, 384-well micotiter plates but also emerging methods such as the nanowell method described by Schullek *et al.* Anal Biochem., 246, 20-29, (1997).

Fusion proteins, such as those made from Fc portion and MAGI polypeptide, as hereinbefore described, can also be used for high-throughput screening assays to identify antagonists for the
35 polypeptide of the present invention (see D. Bennett *et al.*, J Mol Recognition, 8:52-58 (1995); and K. Johanson *et al.*, J Biol Chem., 270(16):9459-9471 (1995)).

The polynucleotides, polypeptides and antibodies to the polypeptide of the present invention may also be used to configure screening methods for detecting the effect of added compounds on the
40 production of mRNA and polypeptide in cells. For example, an ELISA assay may be constructed for measuring secreted or cell associated levels of polypeptide using monoclonal and polyclonal antibodies by standard methods known in the art. This can be used to discover agents that may inhibit or enhance the

production of polypeptide (also called antagonist or agonist, respectively) from suitably manipulated cells or tissues.

A polypeptide of the present invention may be used to identify membrane bound or soluble receptors, if any, through standard receptor binding techniques known in the art. These include, but are not limited to, ligand binding and crosslinking assays in which the polypeptide is labeled with a radioactive isotope (for instance, ^{125}I), chemically modified (for instance, biotinylated), or fused to a peptide sequence suitable for detection or purification, and incubated with a source of the putative receptor (cells, cell membranes, cell supernatants, tissue extracts, bodily fluids). Other methods include biophysical techniques such as surface plasmon resonance and spectroscopy. These screening methods may also be used to identify agonists and antagonists of the polypeptide that compete with the binding of the polypeptide to its receptors, if any. Standard methods for conducting such assays are well understood in the art.

Examples of antagonists of polypeptides of the present invention include antibodies or, in some cases, oligonucleotides or proteins that are closely related to the ligands, substrates, receptors, enzymes, etc., as the case may be, of the polypeptide, e.g., a fragment of the ligands, substrates, receptors, enzymes, etc.; or a small molecule that bind to the polypeptide of the present invention but do not elicit a response, so that the activity of the polypeptide is prevented.

Screening methods may also involve the use of transgenic technology and MAGI gene. The art of constructing transgenic animals is well established. For example, the MAGI gene may be introduced through microinjection into the male pronucleus of fertilized oocytes, retroviral transfer into pre- or post-implantation embryos, or injection of genetically modified, such as by electroporation, embryonic stem cells into host blastocysts. Particularly useful transgenic animals are so-called "knock-in" animals in which an animal gene is replaced by the human equivalent within the genome of that animal. Knock-in transgenic animals are useful in the drug discovery process, for target validation, where the compound is specific for the human target. Other useful transgenic animals are so-called "knock-out" animals in which the expression of the animal ortholog of a polypeptide of the present invention and encoded by an endogenous DNA sequence in a cell is partially or completely annulled. The gene knock-out may be targeted to specific cells or tissues, may occur only in certain cells or tissues as a consequence of the limitations of the technology, or may occur in all, or substantially all, cells in the animal. Transgenic animal technology also offers a whole animal expression-cloning system in which introduced genes are expressed to give large amounts of polypeptides of the present invention.

Screening kits for use in the above described methods form a further aspect of the present invention. Such screening kits comprise:

- (a) a polypeptide of the present invention;
- (b) a recombinant cell expressing a polypeptide of the present invention;
- (c) a cell membrane expressing a polypeptide of the present invention; or
- (d) an antibody to a polypeptide of the present invention;

which polypeptide is preferably that of SEQ ID NO:2 or SEQ ID NO:6.

It will be appreciated that in any such kit, (a), (b), (c) or (d) may comprise a substantial component.

Glossary

The following definitions are provided to facilitate understanding of certain terms used frequently hereinbefore.

"Antibodies" as used herein includes polyclonal and monoclonal antibodies, chimeric, single chain, and humanized antibodies, as well as Fab fragments, including the products of an

5 Fab or other immunoglobulin expression library.

"Isolated" means altered "by the hand of man" from its natural state, i.e., if it occurs in nature, it has been changed or removed from its original environment, or both. For example, a polynucleotide or a polypeptide naturally present in a living organism is not "isolated," but the same polynucleotide or polypeptide separated from the coexisting materials of its natural state is 10 "isolated", as the term is employed herein. Moreover, a polynucleotide or polypeptide that is introduced into an organism by transformation, genetic manipulation or by any other recombinant method is "isolated" even if it is still present in said organism, which organism may be living or non-living.

"Polynucleotide" generally refers to any polyribonucleotide (RNA) or polydeoxyribonucleotide

15 (DNA), which may be unmodified or modified RNA or DNA. "Polynucleotides" include, without limitation, single- and double-stranded DNA, DNA that is a mixture of single- and double-stranded regions, single- and double-stranded RNA, and RNA that is mixture of single- and double-stranded regions, hybrid molecules comprising DNA and RNA that may be single-stranded or, more typically,

20 double-stranded or a mixture of single- and double-stranded regions. In addition, "polynucleotide" refers to triple-stranded regions comprising RNA or DNA or both RNA and DNA. The term "polynucleotide" also includes DNAs or RNAs containing one or more modified bases and DNAs or RNAs with backbones modified for stability or for other reasons. "Modified" bases include, for example, tritylated bases and unusual bases such as inosine. A variety of modifications may be made to DNA and RNA; thus, "polynucleotide" embraces chemically, enzymatically or metabolically modified forms of polynucleotides 25 as typically found in nature, as well as the chemical forms of DNA and RNA characteristic of viruses and cells. "Polynucleotide" also embraces relatively short polynucleotides, often referred to as oligonucleotides.

"Polypeptide" refers to any polypeptide comprising two or more amino acids joined to each other by peptide bonds or modified peptide bonds, i.e., peptide isosteres. "Polypeptide" refers to both short

30 chains, commonly referred to as peptides, oligopeptides or oligomers, and to longer chains, generally referred to as proteins. Polypeptides may contain amino acids other than the 20 gene-encoded amino acids. "Polypeptides" include amino acid sequences modified either by natural processes, such as post-translational processing, or by chemical modification techniques that are well known in the art. Such modifications are well described in basic texts and in more detailed monographs, as well as in a

35 voluminous research literature. Modifications may occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. It will be appreciated that the same type of modification may be present to the same or varying degrees at several sites in a given polypeptide. Also, a given polypeptide may contain many types of modifications. Polypeptides may be branched as a result of ubiquitination, and they may be cyclic, with or without branching. Cyclic, 40 branched and branched cyclic polypeptides may result from post-translation natural processes or may be made by synthetic methods. Modifications include acetylation, acylation, ADP-ribosylation, amidation, biotinylation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment

of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphotidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cystine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation,

5 myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination (see, for instance, Proteins - Structure and Molecular Properties, 2nd Ed.,

T. E. Creighton, W. H. Freeman and Company, New York, 1993; Wold, F., Post-translational Protein Modifications: Perspectives and Prospects, 1-12, in Post-translational Covalent Modification of Proteins.

10 B. C. Johnson, Ed., Academic Press, New York, 1983; Seifter *et al.*, "Analysis for protein modifications and nonprotein cofactors", Meth Enzymol, 182, 626-646, 1990, and Rattan *et al.*, "Protein Synthesis: Post-translational Modifications and Aging", Ann NY Acad Sci, 663, 48-62, 1992).

"Fragment" of a polypeptide sequence refers to a polypeptide sequence that is shorter than the reference sequence but that retains essentially the same biological function or activity as the reference

15 polypeptide. "Fragment" of a polynucleotide sequence refers to a polynucleotide sequence that is shorter than the reference sequence of SEQ ID NO:1 or SEQ ID NO:5.

"Variant" refers to a polynucleotide or polypeptide that differs from a reference polynucleotide or polypeptide, but retains the essential properties thereof. A typical variant of a polynucleotide differs in nucleotide sequence from the reference polynucleotide. Changes in the nucleotide sequence of the variant

20 may or may not alter the amino acid sequence of a polypeptide encoded by the reference polynucleotide. Nucleotide changes may result in amino acid substitutions, additions, deletions, fusions and truncations in the polypeptide encoded by the reference sequence, as discussed below. A typical variant of a polypeptide differs in amino acid sequence from the reference polypeptide. Generally, alterations are limited so that the sequences of the reference polypeptide and the variant are closely similar overall and,

25 in many regions, identical. A variant and reference polypeptide may differ in amino acid sequence by one or more substitutions, insertions, deletions in any combination. A substituted or inserted amino acid residue may or may not be one encoded by the genetic code. Typical conservative substitutions include Gly, Ala; Val, Ile, Leu; Asp, Glu; Asn, Gln; Ser, Thr; Lys, Arg; and Phe and Tyr. A variant of a polynucleotide or polypeptide may be naturally occurring such as an allele, or it may be a variant that is

30 not known to occur naturally. Non-naturally occurring variants of polynucleotides and polypeptides may be made by mutagenesis techniques or by direct synthesis. Also included as variants are polypeptides

having one or more post-translational modifications, for instance glycosylation, phosphorylation,

methylation, ADP ribosylation and the like. Embodiments include methylation of the N-terminal amino acid, phosphorylations of serines and threonines and modification of C-terminal glycines.

35 "Allele" refers to one of two or more alternative forms of a gene occurring at a given locus in the genome.

"Polymorphism" refers to a variation in nucleotide sequence (and encoded polypeptide sequence, if relevant) at a given position in the genome within a population.

40 "Single Nucleotide Polymorphism" (SNP) refers to the occurrence of nucleotide variability at a single nucleotide position in the genome, within a population. An SNP may occur within a gene or within intergenic regions of the genome. SNPs can be assayed using Allele Specific Amplification (ASA). For the process at least 3 primers are required. A common primer is used in reverse complement to the

polymorphism being assayed. This common primer can be between 50 and 1500 bps from the polymorphic base. The other two (or more) primers are identical to each other except that the final 3' base wobbles to match one of the two (or more) alleles that make up the polymorphism. Two (or more) PCR reactions are then conducted on sample DNA, each using the common primer and one of the Allele

5 Specific Primers.

"Splice Variant" as used herein refers to cDNA molecules produced from RNA molecules initially transcribed from the same genomic DNA sequence but which have undergone alternative RNA splicing. Alternative RNA splicing occurs when a primary RNA transcript undergoes splicing, generally for the removal of introns, which results in the production of more than one mRNA molecule each of that may encode different amino acid sequences. The term splice variant also refers to the proteins encoded by the above cDNA molecules.

"Identity" reflects a relationship between two or more polypeptide sequences or two or more polynucleotide sequences, determined by comparing the sequences. In general, identity refers to an exact nucleotide to nucleotide or amino acid to amino acid correspondence of the two polynucleotide or two 10 polypeptide sequences, respectively, over the length of the sequences being compared.

"% Identity" - For sequences where there is not an exact correspondence, a "% identity" may be determined. In general, the two sequences to be compared are aligned to give a maximum correlation between the sequences. This may include inserting "gaps" in either one or both sequences, to enhance the degree of alignment. A % identity may be determined over the whole length of each of the sequences 20 being compared (so-called global alignment), that is particularly suitable for sequences of the same or very similar length, or over shorter, defined lengths (so-called local alignment), that is more suitable for sequences of unequal length.

"Similarity" is a further, more sophisticated measure of the relationship between two polypeptide sequences. In general, "similarity" means a comparison between the amino acids of two polypeptide 25 chains, on a residue by residue basis, taking into account not only exact correspondences between a between pairs of residues, one from each of the sequences being compared (as for identity) but also, where there is not an exact correspondence, whether, on an evolutionary basis, one residue is a likely substitute for the other. This likelihood has an associated "score" from which the "% similarity" of the two sequences can then be determined.

30 Methods for comparing the identity and similarity of two or more sequences are well known in the art. Thus for instance, programs available in the Wisconsin Sequence Analysis Package, version 9.1 (Devereux J et al, Nucleic Acids Res. 12, 387-395, 1984, available from Genetics Computer Group, Madison, Wisconsin, USA), for example the programs BESTFIT and GAP, may be used to determine the % identity between two polynucleotides and the % identity and the % similarity between two polypeptide 35 sequences. BESTFIT uses the "local homology" algorithm of Smith and Waterman (J Mol Biol, 147, 195-197, 1981, Advances in Applied Mathematics, 2, 482-489, 1981) and finds the best single region of similarity between two sequences. BESTFIT is more suited to comparing two polynucleotide or two polypeptide sequences that are dissimilar in length, the program assuming that the shorter sequence represents a portion of the longer. In comparison, GAP aligns two sequences, finding a "maximum 40 similarity", according to the algorithm of Neddleman and Wunsch (J Mol Biol, 48, 443-453, 1970). GAP is more suited to comparing sequences that are approximately the same length and an alignment is expected over the entire length. Preferably, the parameters "Gap Weight" and "Length Weight" used in

each program are 50 and 3. for polynucleotide sequences and 12 and 4 for polypeptide sequences, respectively. Preferably, % identities and similarities are determined when the two sequences being compared are optimally aligned.

Other programs for determining identity and/or similarity between sequences are also known in the art, for instance the BLAST family of programs (Altschul S F et al, J Mol Biol, 215, 403-410, 1990, Altschul S F et al, Nucleic Acids Res., 25:389-3402, 1997, available from the National Center for Biotechnology Information (NCBI), Bethesda, Maryland, USA and accessible through the home page of the NCBI at www.ncbi.nlm.nih.gov) and FASTA (Pearson W R, Methods in Enzymology, 183, 63-99, 1990; Pearson W R and Lipman D J, Proc Nat Acad Sci USA, 85, 2444-2448, 1988, available as part of the Wisconsin Sequence Analysis Package).

Preferably, the BLOSUM62 amino acid substitution matrix (Henikoff S and Henikoff J G, Proc. Nat. Acad. Sci. USA, 89, 10915-10919, 1992) is used in polypeptide sequence comparisons including where nucleotide sequences are first translated into amino acid sequences before comparison.

Preferably, the program BESTFIT is used to determine the % identity of a query polynucleotide or a polypeptide sequence with respect to a reference polynucleotide or a polypeptide sequence, the query and the reference sequence being optimally aligned and the parameters of the program set at the default value, as hereinbefore described.

"Identity Index" is a measure of sequence relatedness which may be used to compare a candidate sequence (polynucleotide or polypeptide) and a reference sequence. Thus, for instance, a candidate polynucleotide sequence having, for example, an Identity Index of 0.95 compared to a reference polynucleotide sequence is identical to the reference sequence except that the candidate polynucleotide sequence may include on average up to five differences per each 100 nucleotides of the reference sequence. Such differences are selected from the group consisting of at least one nucleotide deletion, substitution, including transition and transversion, or insertion. These differences may occur at the 5' or 20 3' terminal positions of the reference polynucleotide sequence or anywhere between these terminal positions, interspersed either individually among the nucleotides in the reference sequence or in one or more contiguous groups within the reference sequence. In other words, to obtain a polynucleotide sequence having an Identity Index of 0.95 compared to a reference polynucleotide sequence, an average of up to 5 in every 100 of the nucleotides of the reference sequence may be deleted, substituted or 25 inserted, or any combination thereof, as hereinbefore described. The same applies *mutatis mutandis* for 30 other values of the Identity Index, for instance 0.96, 0.97, 0.98 and 0.99.

Similarly, for a polypeptide, a candidate polypeptide sequence having, for example, an Identity Index of 0.95 compared to a reference polypeptide sequence is identical to the reference sequence except that the polypeptide sequence may include an average of up to five differences per each 100 amino acids 35 of the reference sequence. Such differences are selected from the group consisting of at least one amino acid deletion, substitution, including conservative and non-conservative substitution, or insertion. These differences may occur at the amino- or carboxy-terminal positions of the reference polypeptide sequence or anywhere between these terminal positions, interspersed either individually among the amino acids in the reference sequence or in one or more contiguous groups within the reference sequence. In other 40 words, to obtain a polypeptide sequence having an Identity Index of 0.95 compared to a reference polypeptide sequence, an average of up to 5 in every 100 of the amino acids in the reference sequence

may be deleted, substituted or inserted, or any combination thereof, as hereinbefore described. The same applies *mutatis mutandis* for other values of the Identity Index, for instance 0.96, 0.97, 0.98 and 0.99.

The relationship between the number of nucleotide or amino acid differences and the Identity Index may be expressed in the following equation:

5

$$n_a \leq x_a - (x_a \cdot I),$$

in which:

n_a is the number of nucleotide or amino acid differences.

x_a is the total number of nucleotides or amino acids in SEQ ID NO:1 or SEQ ID NO:5 or SEQ ID NO:2 or SEQ ID NO:6, respectively,

10 I is the Identity Index .

• is the symbol for the multiplication operator, and

in which any non-integer product of x_a and I is rounded down to the nearest integer prior to subtracting it from x_a .

"Homolog" is a generic term used in the art to indicate a polynucleotide or polypeptide sequence possessing a high degree of sequence relatedness to a reference sequence. Such relatedness may be quantified by determining the degree of identity and/or similarity between the two sequences as hereinbefore defined. Falling within this generic term are the terms "ortholog", and "paralog".

"Ortholog" refers to a polynucleotide or polypeptide that is the functional equivalent of the polynucleotide or polypeptide in another species. "Paralog" refers to a polynucleotide or polypeptide that within the same species which is functionally similar.

"Fusion protein" refers to a protein encoded by two, often unrelated, fused genes or fragments thereof. In one example, EP-A-0 464 *** discloses fusion proteins comprising various portions of constant region of immunoglobulin molecules together with another human protein or part thereof. In many cases, employing an immunoglobulin Fc region as a part of a fusion protein is advantageous for use in therapy and diagnosis resulting in, for example, improved pharmacokinetic properties [see, e.g., EP-A 0232 262]. On the other hand, for some uses it would be desirable to be able to delete the Fc part after the fusion protein has been expressed, detected and purified.

Example – Tissue distribution of human MAGI

A radioactively labelled DNA probe corresponding to the C-terminal 2kb of the MAGI open reading frame was hybridised with a Clontech MTN-1 filter containing equivalent loadings of mRNA from each of the indicated tissues and then washed to high stringency to only detect MAGI transcripts. The results are shown in figure 1. At least three bands are visible (>5kb, 2.4kb and <2kb). The largest transcript is present in adult brain and to a lesser degree in heart and skeletal muscle. The middle transcript is present at essentially similar levels in all tissues while the shortest transcripts are more specifically expressed, most abundantly in skeletal muscle, brain and kidney with low levels detectable in the pancreas.

Figure 1 Depicts a Northern blot showing tissue distribution of human MAGI. Lane 1 contains human adult heart RNA, lane 2 contains human adult brain RNA, lane 3 contains human placental RNA, lane 4 contains human adult lung RNA, lane 5 contains human adult liver RNA, lane 6 contains human adult skeletal muscle RNA, lane 7 contains human adult kidney RNA and lane 8 contains human adult pancreas RNA.

All publications and references, including but not limited to patents and patent applications, cited in this specification are herein incorporated by reference in their entirety as if each individual publication or reference were specifically and individually indicated to be incorporated by reference herein as being fully 5 set forth. Any patent application to which this application claims priority is also incorporated by reference herein in its entirety in the manner described above for publications and references.

SEQUENCE INFORMATION

SEQ ID NO:1

ATGGAAGACCTGGACCAGTCTCCTCTGGTCTCGTCCTCGAACAGCCCACCCCGGCCAG
5 CCCGC GTTCAAGTACCACTAGCTGAGGGAGCCCCAGGACGAGGAGGAAGAAGAGGAGGAG
GAAGAGGAGGACGAGGACGAAGACCTGGAGGAGCTGGAGGTCTGGAGAGGAAGCCGCC
GCCGGGCTGTCCCGGGCCCCAGTGCCCCACCGCCCCCTGCCGCCGGCGGCCCTGATGGAC
TTCGGAAATGACTTCGTGCCGGCGCCCCGGGGACCCCTGCCGGCGCTCCCCCGTC
GCCCGGAGCGGCAGCGTCTGGGACCCGAGCCGGTGTCTGACCGTGCCCGCCA
10 TCCCCGCTGTCTGCCGCAGTCTGCCCTCCAAGCTCCCTGAGGACGACGAGCCTCCG
GCCCGGCTCCCCCTCTCCCCGGCAGCGTGA GCCCCCAGGCAGAGCCGTGTGGACC
CCGCCAGCCCCGGCTCCCGCCGCCCGCCCCCTCCACCCGGCGGCCAACGCCAGGGGC
TCCTCGGGCTCAGTGGATGAGACCTTTTGCTTCTCTGCTGCATCTGAGCCTGTGATA
CGCTCCTCTGCA GAAAATATGGACTTGAAGGAGCAGCCAGGTAAACACTATTCGGCTGGT
15 CAAGAGGATTCCCCATCTGCTCTGAAACTGCTGCTCTCTCTCTGCTCCT
CTCTCAGCCGCTTCTTCAAAGAACATGAATACCTTGGTAATTGTCAACAGTATTACCC
ACTGAAGGAACACTTCAAGAAAATGTCAAGTGAAGCTCTAAAGAGGTCTAGAGAAGGCA
AAAACTCTACTCATAGATAGAGATTTAACAGAGTTTCAAGATTAGAATAACTCAGAAATG
GGATCATCGTTCA GTGTCTCTCCAAAAGCAGAACATCTGCCGTAA TAGTAGCAAATCCTAGG
20 GAAGAAATAATCGTAAAAAATAAAGATGAAGAAGAGTTAGTTAGTAATAACATCCTT
CATATCAACAAGAGTTACCTACAGCTCTACTAAATTGGTTAAAGAGGATGAAGTTGTG
TCTTCAGAAAAAGCAAAGACAGTTTAATGAAAAGAGAGTTGCAAGTGGCTCTATG
AGGGAGGAATATGCAGACTTCAAACCATTGAGCGAGTATGGGAAGTGAAGATACTAAG
GAAGATAGTGATATGTTGGCTGGAGGTAAATCGAGAGCAACTTGAAAGTAAAGTG
25 GATAAAAATGTTGCAGATAGCCTTGAGCAA ACTAATCACGAAAAGATAGTGAGAGT
AGTAATGATGATACTTCTTCCCCAGTACGCCAGAACGGTATAAAGGATCGTCCAGGAGCA
TATATCACATGTGCTCCCTTAACCCAGCAGCAACTGAGAGCATTGCAACAAACATTTT
CCTTGTTAGGAGATCCTACTTCAGAAAATAAGACCGATGAAAAAAAATAGAAGAAAAG
AAGGCCAAATAGTAACAGAGAAGAACTAGCACCAAAACATCAAACCCCTTTCTTGT
30 GCAGCACAGGATTCTGAGACAGATTATGTACAACAGATAATTAAACAAGGTGACTGAG
GAAGTCGTGGCAAACATGCTGAAGGCTGACTCCAGATTAGTACAGGAAGCATGTGAA
AGTGAATTGAATGAAGTTACTGGTACAAAGATTGCTTATGAAACAAAAATGGACTTGGTT
CAAACATCAGAAGTTATGCAAGAGTCACTCTATCCTGCAGCACAGCTTGCCCATCATTT
GAAGAGTCAGAAGCTACTCCTCACCAAGTTTGCTGACATTGTTATGGAAGCACCATTG
35 AATTCTGCAGTCTCTAGTGTGGTCTCCGTGATACAGCCCAGCTCATCACCATTAGAA
GCTTCTTCAGTTAATTATGAAAGCATAAAACATGAGCCTGAAAACCCCCCACCATATGAA
GAGGCCATGAGTGTATCACTAAAAAGTATCAGGAATAAGGAAGAAATTAAAGAGCCT
GAAAATTTAATGCAGCTCTCAAGAACAGAACGCTCCTTATATATCTATTGCATGTGAT
TTAATTAAAGAAACAAAGCTTCTGCTGAACCAGCTCCGGATTCTCTGATTATTCA
40 ATGGCAAAAGTTGAACAGCCAGTGCCTGATCATTCTGAGCTAGTTGAAGATTCTCACCT
GATTCTGAACCAGTTGACTTATTTAGTGTGATTCAATACCTGACGTTCCACAAAACAA
GATGAAAATGTGATGCTTGAAAGAAAGTCTCACTGAGACTTCATTGAGTCATGATA

GAATATGAAAATAAGGAAAAACTCAGTGCTTGCCACCTGAGGGAGGAAAGCCATATTG
 GAATCTTTAAGCTCAGTTAGATAACACAAAAGATAACCCTGTTACCTGATGAAGTTCA
 ACATTGAGCAAAAGGAGAAAATTCCCTTGCAGATGGAGGAGCTCAGTACTGCAGTTAT
 TCAAATGATGACTTATTTATTCTAAGGAAGCACAGATAAGAGAAACTGAAACGTTCA
 5 GATTCATCTCCAATTGAAATTATAGATGAGTCCCTACATTGATCAGTTCTAAAACGTGAT
 TCATTTCTAAATTAGCCAGGAAATATACTGACCTAGAAGTATCCCACAAAAGTGAATT
 GCTAATGCCCGGATGGAGCTGGTCATTGCCTGCACAGAATTGCCCATGACCTTCT
 TTGAAGAACATACAACCCAAAGTTGAAGAGAAATCAGTTCTCAGATGACTTTCTAAA
 AATGGGTCTGCTACATCAAAGGTGCTTATTGCCTCCAGATGTTCTGCTTGGCCACT
 10 CAAGCAGAGATAGAGAGCATAGTTAAACCCAAAGTTCTGTGAAAGAAGCTGAGAAAAAA
 CTTCCCTCCGATACAGAAAAAGAGGACAGATCACCCTGCTATATTTCAGCAGAGCTG
 AGTAAAACCTTCAGTTGTTGACCTCCTGACTGGAGAGACATTAAGAAGACTGGAGTGGTG
 TTTGGTCCAGCCTATTCTGCTGCTTCATTGACAGTATTGAGCATTGAGCGTAACA
 GCCTACATTGCCTGGCCCTGCTCTGTGACCATCAGCTTAGGATATACAAGGGTGTG
 15 ATCCAAGCTATCCAGAAATCAGATGAAGGCCACCCATTCAAGGCATATCTGAATCTGAA
 GTTGCTATATCTGAGGAGTTGGTTCAGAAGTACAGTAATTCTGCTCTGGTCATGTGAAC
 TGCACGATAAAGGAACTCAGGCGCCTCTCTAGTTGATGATTGATTCTCTGAAG
 TTTGCAGTGTGATGTGGTATTTACCTATGTTGTCCTGTTAATGGTCTGACACTA
 CTGATTTGGCTCTCATTCACTCTCAGTGTGCTGTTAATGGTCTGACACTA
 20 CAGATAGATCATTATCTAGGACTTGCAAATAAGAATGTTAAAGATGCTATGGCTAAAATC
 CAAGCAAAATCCCTGGATTGAAGCGCAAAGCTGAATGA

SEQ ID NO:2

MEDLDQSPLVSSSDSPPRPQPAFKYQFVREPEDEEEEEEEEEEDEDLELEVLERKPA
 25 AGLSAAPVPTAPAAGAPLMDFGNDFVPPAPRGPLAAPPVAPERQPSWDPSPVSSTVPAP
 SPLSAAAAPSPLPEDDEPPARPPPPPASVSPQAEPVWTPPAPAPAAPPSTPAPKRRG
 SSGSVDETLFALPAASEPVISSAENMDLKEQPGNTISAGQEDFPSVLLETAASLPSLSP
 LSAASFKEHEYLGNLSTVLPTEGTLQENVSEASKEVSEAKTLLIDRDLTEFSELEYSEM
 GSSFSVSPKAESAVIVANPREEITVKNKDEEEKLVSNNILHNQQELPTALTKLVKEDEVV
 30 SSEKAKDSFNEKRVAVEAPMREEYADFKPFERVWEVKDSKEDSDMLAAGGKIESNLESKV
 DKKCFADSLEQTNHEDSESSNDDTSFPSTPEGIKDRPGAYITCAPFNPATESIATNIF
 PLLGDPTESENKTDEKKIEEKKAQIVTEKNTSTKTSNPFVAAQDSETDYVTTDNLTkvTE
 EVVANMPGLTPDLVQEACESELNEVTGKIAVETKMDLVQTSEVMQESLYPAAQLCPSF
 ESEATPSPVLPDIVMEAPLNSAVPSAGASVIQPSSPLEASSVNYESIKHEPENPPPYE
 35 EAMSVSLKKVSGIKEEIKEPENINAALQETEAPYISIACDLIKETKLSAEPAPDFSDYSE
 MAKVEQPVPDHSELVEDSSPDSEPVDLFSDDSIPDVPQKQDETVMVLVKESLTETSFESMI
 EYENKEKLSALPPEGGKPYLESFKLSLDNTKDTLLPDEVSTLSKKEKIPLQMEELSTAVY
 SNDDLFISKEAQIRETETFSDDSSPIEIIDEFPTLISSKTDFSKLAREYTDLEVSHKSEI
 ANAPDGAGSLPCTELPHDLSLKNIQPKVEEKISFSDDFSKNGSATSKVLLPPDVSALAT
 40 QAEIESIVKPKVLVKEAEKKLPSDTEKEDRSPSAIFSAELSKTSVVDLLYWRDIKKTGVV
 FGASLFLLLSLTVFSIVSVTAYIALALLSVTISFRYKGVIQAIQKSDEGHPFRAYLESE
 VAISEELVQKYSNSALGHVNCTIKELRRLFLVDDLVDSLKFAVLMWVFTYVGALFNGLTL

LILALISLFSVPVIYERHQAQIDHYLGLANKNVKDAMAKIQAKIPGLKRKAE

SEQ ID NO:3

GAAAATATGGACTTGAAGGAGCAGCCAGGTAAACACTATTCGGCTGGTCAAGAGGATTTC
 5 CCATCTGCTCTGTTGAAACTGCTGCTCTNTTCCTCTGTCTCCTCTCAGCCGCT
 TCTTCAAAGAACATGAATACCTTGGTAATTGTCAACAGTATTACCACTGAAGGAACA
 CTTCAAGAAAATGTCAGTGAAGCTCTAAAGAGGTCTCAGAGAAGGAAAAACTCTACTC
 ATAGATAGAGATTTAACAGAGTTTCAAATTAGAATACTCAGAAATGGGATCATCGTTC
 AGTGTCTCTC AAAAGCAGAATCTGCCGTAAAGTAGCAAATCCTAGGAAAGAAATAATC
 10 GTGAAAATAAGATGAAGAAGAGAAGTTAGTTAGTAATAACATCCTCATANTCAACAA
 GAGTTACCTACAGCTCTACTAAATTGGTTAAAGAGGATGAAGTTGTCTTCAGAAAAAA
 GCAGAAAGACAGTTTAATGAAAAGAGAGTTGCAGTGGAAAGCTCTATGAGGGAGGAATAT
 GCAGACTTCAAACCATTGAGCGAGTATGGAAAGTAAAGATAGTAAGGAAGATAGTGT
 ATGTTGGCTGCTGGAGGTAAATCAGAGAGCAACTGGAAAGTAAAGTGGATAAAAATGT
 15 TTTGCAGATAGCCTTGAGCAAACTAATCACGAAAAGATAGTGTAGAGTAGTAATGATGAT
 ACTTCTTCCCCAGTACGCCAGAAGGTATAAAGGATGTTCAAGGAGCATATATCACATGT
 GCTCCCTTAACCCAGCAGCAACTGAGAGCATTGCAACAAACATTTCTTTGTTAGGA
 GATCCTACTTCAGAAAATAAGACCGATG

20 SEQ ID NO:4

ENMDLKEQPGNTISAGQEDFPSVLLETAASXPSSLPLSAASFKEHEYLGNLSTVLPTEGT
 LQENVSEASKEVSEKAKTLLIDRDLTEFSELEYSEMSSFSVSPKAESAVIVANPREEII
 VKNKDEEEKLVSNNILHXQQELPTALTALKVKEDEVVSSEKAKDSFNEKRVAVEAPMREYY
 ADFKPFERVWEVKSKEDSDMLAAGGKIESNLESKVDKKCFADSLEQTNHEKDSESSNDD
 25 TSFPSTPEGIKDRSGAYITCAPFNPAATESIATNIFPLLGDPSENKTD

SEQ ID NO:5

ATGGAAGACCTGGACCAGTCTCCTCTGGTCTCGCCTCGGACAGCCCACCCGGCCGAG
 CCCCGTTCAAGTACCACTGGAGGGAGCCCGAGGACGAGGAGGAAGAAGAGGAGGAG
 30 GAAGAGGAGGACGAGGACGAAGACCTGGAGGAGCTGGAGGTGCTGGAGAGGAAGCCC
 GCCGGGCTGTCCCGGGCCCCAGTGCCACCAGCCCTGCCGGCGCGCCCTGATGGAC
 TTCGAAATGACTTCGTGCCGCCGGGGACCCCTGCCGGCGCTCCCCCGTC
 GCCCGGAGCGCAGCCGTCTGGGACCCGAGCCGGTGTCTGACCGTGCCCGCCA
 TCCCGCTGTCTGCCAGTCTGCCCTCAAGCTCCCTGAGGACGAGCAGCTCCG
 35 GCCCGGCTCCCCCTCTCCCCGGCCAGCGTGAGCCCCCAGGCAGAGCCGTGTGGACC
 CCGCCAGCCCCGGCTCCCGCCGCCCTCCACCCGGCGGCCAAGCGCAGGGGC
 TCCTCGGGCTCAGTGGTTGTTGACCTCTGTACTGGAGAGACATTAAGAAGACTGGAGT
 GTGTTGGTGCCAGCCTATTCTGCTGCTTCAATTGACAGTATTGAGCATTGTGAGCGTA
 ACAGCCTACATTGCCCTGGCCCTGCTCTGTGACCATCAGCTTCTGAGGATATA
 40 GTGATCCAAGCTATCCAGAAATCAGATGAAGGCCACCCATTGAGGCATATCTGGAAATCT
 GAAGTTGCTATATCTGAGGAGTTGGTTCAAGAAGTACAGTAATTCTGCTTGGTCATGT
 AACTGCACGATAAAGGAACCTCAGGCCCTTCTAGTTGATGATTGATTCTCTG

AAGTTTGCAGTGGATGTGGGATTACCTATGTTGGCCTGTTAATGGCTGACA
CTACTGATTTGGCTCTCATTCACTTCAGTGGCCTGTTATTTATGAACGGCATCAG
GCACAGATAGATCATTATCTAGGACTTCAAATAAGAATGTTAAAGATGCTATGGCTAAA
ATCCAAGCAAAATCCCTGGATTGAAGCGAAAGCTGAATGA

5

SEQ ID NO:6

MEDLDQSPLVSSSDSPPRPQPAFKYQFVREPEDEEEEEEEEEEDEDLEELEVLERKPA
AGLSAAPVPTAPAAGAPLMDFGNDVPPAPRGPLPAAPPVAPERQPSWDPSPVSSTVPAP

SPLSAAAVSPSKLPEDDEPPARPPPPPPASVSPQAEPVWTPPAPAPAAPSTPAAPKRRG

10 SSGSVVVVDLLYWRDIKKTGVVFGASLFLLSLTVFSIVSVTAYIALALLSVTISFRIYKG
VIQAIQKSDEGHPFRAYLESEVAISEELVQKYSNSALGHVNCTIKELRRFLVDDLVDSL
KFAVLMWVFTYVGALFNGLTLLILALISLFSVPVIYERHQAQIDHYLGLANKNVKDAMAK
IQAKIPGLKRKAE

Claims

1. An isolated polypeptide selected from the group consisting of:

- (a) an isolated polypeptide encoded by a polynucleotide comprising the sequence of SEQ ID NO:1 or SEQ ID NO:5;
- (b) an isolated polypeptide comprising a polypeptide sequence having at least 95% identity to the polypeptide sequence of SEQ ID NO:2 or SEQ ID NO:6;
- (c) an isolated polypeptide having at least 95% identity to the polypeptide sequence of SEQ ID NO:2 or SEQ ID NO:6; and
- (d) fragments and variants of such polypeptides in (a) to (e).

5 2. The isolated polypeptide as claimed in claim 1 comprising the polypeptide sequence of SEQ ID NO:2 or SEQ ID NO:6.

10 3. The isolated polypeptide as claimed in claim 1 which is the polypeptide sequence of SEQ ID NO:2 or SEQ ID NO:6.

15 4. An isolated polynucleotide selected from the group consisting of:

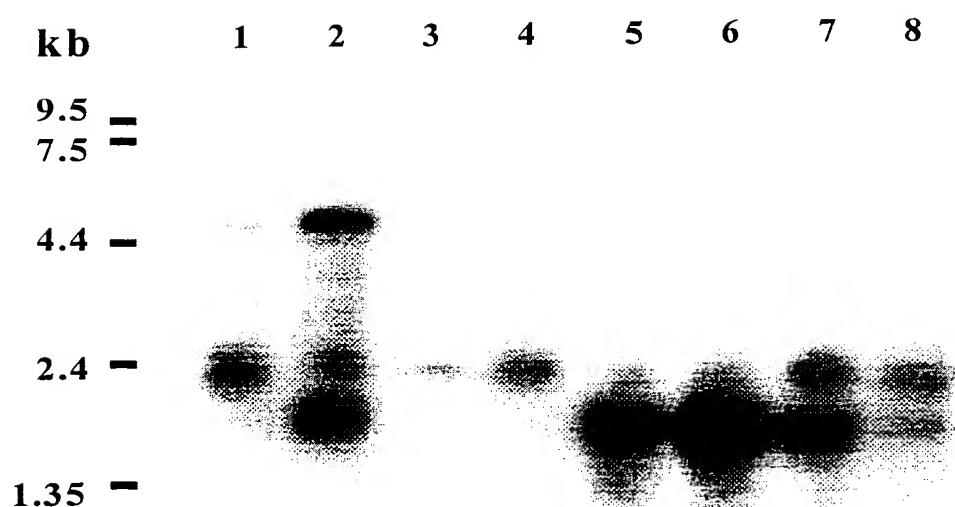
- (a) an isolated polynucleotide comprising a polynucleotide sequence having at least 95% identity to the polynucleotide sequence of SEQ ID NO:1 or SEQ ID NO:5;
- (b) an isolated polynucleotide having at least 95% identity to the polynucleotide of SEQ ID NO:1 or SEQ ID NO:5;
- (c) an isolated polynucleotide comprising a polynucleotide sequence encoding a polypeptide sequence having at least 95% identity to the polypeptide sequence of SEQ ID NO:2 or SEQ ID NO:6;
- (d) an isolated polynucleotide having a polynucleotide sequence encoding a polypeptide sequence having at least 95% identity to the polypeptide sequence of SEQ ID NO:2 or SEQ ID NO:6;
- (e) an isolated polynucleotide with a nucleotide sequence of at least 100 nucleotides obtained by screening a library under stringent hybridization conditions with a labeled probe having the sequence of SEQ ID NO: 1 or SEQ ID NO:5 or a fragment thereof having at least 15 nucleotides;
- (f) a polynucleotide which is the RNA equivalent of a polynucleotide of (a) to (e);
or a polynucleotide sequence complementary to said isolated polynucleotide
and polynucleotides that are variants and fragments of the above mentioned polynucleotides or that are complementary to above mentioned polynucleotides, over the entire length thereof.

20 5. An isolated polynucleotide as claimed in claim 4 selected from the group consisting of:

- (a) an isolated polynucleotide comprising the polynucleotide of SEQ ID NO:1 or SEQ ID NO:5;
- (b) the isolated polynucleotide of SEQ ID NO:1 or SEQ ID NO:5;
- (c) an isolated polynucleotide comprising a polynucleotide sequence encoding the polypeptide of SEQ ID NO:2 or SEQ ID NO:6; and
- (d) an isolated polynucleotide encoding the polypeptide of SEQ ID NO:2 or SEQ ID NO:6.

6. An expression system comprising a polynucleotide capable of producing a polypeptide of claim 1 when said expression vector is present in a compatible host cell.
7. A recombinant host cell comprising the expression vector of claim 6 or a membrane thereof expressing the polypeptide of claim 1.
5
8. A process for producing a polypeptide of claim 1 comprising the step of culturing a host cell as defined in claim 7 under conditions sufficient for the production of said polypeptide and recovering the polypeptide from the culture medium.
10
9. An antibody immunospecific for the polypeptide of any one of claims 1 to 3.
10. A method for screening to identify compounds that stimulate or inhibit the function or level of the polypeptide of claim 1 comprising a method selected from the group consisting of:
15
- (a) measuring or, detecting, quantitatively or qualitatively, the binding of a candidate compound to the polypeptide (or to the cells or membranes expressing the polypeptide) or a fusion protein thereof by means of a label directly or indirectly associated with the candidate compound;
- (b) measuring the competition of binding of a candidate compound to the polypeptide (or to the cells or membranes expressing the polypeptide) or a fusion protein thereof in the presence of a labeled
20 competitor;
- (c) testing whether the candidate compound results in a signal generated by activation or inhibition of the polypeptide, using detection systems appropriate to the cells or cell membranes expressing the polypeptide;
- (d) mixing a candidate compound with a solution containing a polypeptide of claim 1, to form a mixture,
25 measuring activity of the polypeptide in the mixture, and comparing the activity of the mixture to a control mixture which contains no candidate compound; or
- (e) detecting the effect of a candidate compound on the production of mRNA encoding said polypeptide or said polypeptide in cells, using for instance, an ELISA assay.
- 30 11. An isolated polynucleotide selected from the group consisting of:
 - (a) an isolated polynucleotide comprising a nucleotide sequence which has at least 95% identity to SEQ ID NO:3 over the entire length of SEQ ID NO:3;
 - (b) an isolated polynucleotide comprising the polynucleotide of SEQ ID NO:3; or
 - (c) the polynucleotide of SEQ ID NO:3.
- 35 (d) an isolated polynucleotide comprising a nucleotide sequence encoding a polypeptide which has at least 95% identity to the amino acid sequence of SEQ ID NO:4, over the entire length of SEQ ID NO:4.
12. A polypeptide selected from the group consisting of:
 - (a) a polypeptide which comprises an amino acid sequence which has at least 95% identity to that of SEQ ID NO:4 over the entire length of SEQ ID NO:4;
 - (b) a polypeptide in which the amino acid sequence has at least 95% identity to the amino acid sequence of SEQ ID NO:4 over the entire length of SEQ ID NO:4;
40

- (c) a polypeptide which comprises the amino acid of SEQ ID NO:4;
- (d) a polypeptide which is the polypeptide of SEQ ID NO:4; or
- (e) a polypeptide which is encoded by a polynucleotide comprising the sequence contained in SEQ ID NO:3.

Figure 1

SEQUENCE LISTING

5 <110> SmithKline Beecham plc
10 <120> Novel compounds

15 <130> GP30165
20 <160> 6

25 <170> FastSEQ for Windows Version 3.0

30 <210> 1
 <211> 3579
 <212> DNA
 <213> Homo sapiens

35 <400> 1
40 atggaagacc tggaccagtc tcctctggtc tcgtcctcg 60
 cccgcgttca agtaccagtt cgtgaggagg cccgaggacg 120
 aggaggaaga agaggaggag gaagaggagg
 gaagaggagg acgaggacga agacctggag gagctggagg tgctggagag 180
 gaagcccccc gcccggctgt ccgcggccccc 240
 agtgcaccacc gcccctgcgg cccgcgcgcgc cctgtatggac
45 ttcggaaatg acttcgtgcc gceggcgcccc cggggacccc 300
 tgccggccgc tccccccgtc gccccggagc ggcagccgtc 360
 ttggggacccg agcccggtgt cgtcgaccgt gcccgcgcca
 tcccccgctgt ctgtcgccgc agtctcgccc tccaagetcc 420
 ctgaggacga cgagccctccg gccccgcctc cccctccccc 480
 ccggccagcccc cggctccccc cccggccagc gtgagccccc 540
 aggcagagcc cgtgtggacc ccgcggccccc cgcgcaggggc
50 tccatcggtc cagtggtatga gaccctttt getcttcctg 600
 ctgcattctg cagaaaatat ggacttgaag gagcagccag gtaacactat 660
 ttcggcttgt caagaggatt tcccattgt cctgttgc 720
 aaactgttgc actgtctgtt ctcttccttc tctgtctcc
 ctctcagccg ctttttcaa agaacatgaa taccttggta 780
 atttgtcaac agtattaccc actgaaggaa cacttcaaga 840
 aaatgtcagt gaagtttcta aagaggtctc agagaaggca
55 aaaactctac tcatagatag agattnaaca gagttttcag 900
 aattagaata ctcagaaatg ggatcatctg tcagtgtctc 960
 tccaaaagca gaatctgcgc taatagtagc aaatcctagg
 gaagaataa tcgtaaaaaa taaagatgaa gaagagaagt 1020
 tagtttagtaa taacatccct cataatcaac aagagttacc
 tacagctttt actaaattgg ttaaagagga tgaagttgtg
 tcttcagaaa aagcaaaaaga cagtttaat gaaaagagag 1080
 ttgcagtgaa agctccatg agggaggaat atgcagactt 1140
 caaaccattt gagcagttt gggaaagtgaa agatagtaag
 gaagatagtg atatgttggc tgctggaggt aaaatcgaga 1200
 gcaacttggaa aagtaaagtg gataaaaaat gtttgcaga 1260
 tagccttgag caaactaatc acgaaaaaga tagtgagagt 1320

	agtaatgatg atacttctt ccccagtacg ccagaaggta taaaggatcg tccaggagca	1380
	tatacacat gtgctccctt taaccaggca gcaactgaga gcattgcaac aaacattttt	1440
	ccttgttag gagatcctac ttccagaaaaat aagaccgatg aaaaaaaaaat agaagaaaaag	1500
5	aaggccccaa tagtaacaga gaagaatact agcaccaaaa catcaaacc ttttcttgta	1560
	gcagcacagg attctgagac agattatgtc acaacagata atttaacaaa ggtgactgag	1620
	gaagtctgtgg caaacatgcc tgaaggccctg actccagatt tagtacagga agcatgtgaa	1680
	agtgaattga atgaagttac tggtacaaag attgctttagt aaacaaaaat ggacttggtt	1740
	caaacatcag aagttatgca agagtcaactc tattctgcag cacagcttt cccatcattt	1800
10	gaagagtcag aagctactcc ttccaccagtt ttgcctgaca ttgttatgga agcaccattt	1860
	aattctgcag ttccctagtgc tggtgcttcc gtgatacagc ccagctcatc accattagaa	1920
	gcttcttcag ttaattatga aagcataaaa catgagccctg aaaacccccc accatatgaa	1980
	gaggccatga gtgtatcact aaaaaaagta tcaggaataa aggaagaaaat taaagagcct	2040
	gaaaatatta atgcagctct tcaagaaaaca gaagctccctt atatatctat tgcattgtat	2100
	ttaattaaag aaacaaagct ttctgctgaa ccagctccgg atttctctgaa ttattcagaa	2160
15	atggcaaaag ttgaacagcc agtgcctgat cattctgagc tagttgaaga ttccctcacct	2220
	gattctgaac cagttgactt atttagtgat gattcaatac ctgacgcttcc aaaaaaaca	2280
	gatgaaaactg tgatgcttgc gaaagaaaat ctcactgaga cttcatttga gtcaatgata	2340
	gaatatgaaa ataaggaaaaa actcagtgct ttgcacccctg agggaggaaaa gccatattt	2400
	gaatcttttta agctcagttt agataacaca aaagataaccc tgttacctgta tgaagtttca	2460
20	acattgagca aaaaggagaa aattcccttgc cagatggagg agctcagtgat tgcagtttat	2520
	tcaaatgatg acttattttat ttcttaaggaa gcacagataa gagaaaactgaa aacgttttca	2580
	gattccatctc caatttggaaat tatagatgag ttccctacat tgatcagttc taaaactgat	2640
	tcattttcta aattagccag ggaatataact gacctagaag tatcccacaaa aagtggaaatt	2700
	gctaatgccc cggatggagc tgggtcatttgc cttgcacag aattgccccca tgacctttct	2760
25	tttgaagaaca tacaacccaa agttgaagag aaaatcagtt tctcagatgat cttttctaaa	2820
	aatgggtctg ctacatcaaa ggtgccttta ttgcctccag atgtttctgc tttggccact	2880
	caagcagaga tagagagcat agttaaaccc aaagtttttgc tgaaagaagc tgagaaaaaa	2940
	cttcccttccg atacagaaaaa agaggacaga tcaccatctg ctatattttc agcagagctg	3000
	agtaaaaactt cagtttgtga cttccctgtac tggagagaca ttaagaagac tggagttgg	3060
30	tttgggtgcca gcctatttcttca gctgcatttca ttgcacagtat tcaagcatttgc gagcgtaaaca	3120
	gcctacatttgc cttggccctt gctctctgttgc accatcagct ttaggatata caaggggtgt	3180
	atccaaagctt tccagaaaatc agatgaaggc caccattca gggcatatctt ggaatctgaa	3240
	gttgcataat ctgaggagtt gttcagaag tacagtaatt ctgcattttgc tcatgtgaaac	3300
	tgcacgataa aggaacttcg ggccttttgc tttagttgtat atttagtttgc ttctctgaaag	3360
35	tttgcagtgt tgatgtgggt atttacctat gttgggtcct tttttatgg tctgacacta	3420
	ctgattttgg ctctcatttca acttttcaactt gttccctgtta tttatgaacg gcatcaggcg	3480
	cagatagatc attatctagg acttgcaaaat aagaatgtt aagatgtat ggctaaaatc	3540
	caagccccaa tccctggatt gaagcgcaaa gctgaatga	3579
40	<210> 2	
	<211> 1192	
	<212> PRT	

INTERNATIONAL SEARCH REPORT

Inte 'national Application No
PC /GB 99/02360

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12N15/12 C07K14/47 C07K16/18 G01N33/68 C12Q1/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12N C07K G01N C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>WO 98 06841 A (INCYTE PHARMACEUTICALS INC (US) BANDMAN O; AU-YOUNG J; GOLI S; HILLMAN) 19 February 1998 (1998-02-19) abstract page 2, line 43 -page 3, line 20; figures 2,6 page 7, line 41 -page 9, line 14 page 16, line 23-30 page 18, line 33 -page 19, line 13 page 35, line 33 -page 36, line 7 99% identity between Seq.ID:1 nt.108-674 and Seq.ID:1 of current application nt.3013-3579 99% identity between Seq.ID:2 and Seq.ID:2 of current application aa.1005-1192 page 38 -page 39 page 46; claims 1-14 --- -/-</p>	1-12

Further documents are listed in the continuation of box C.

Patent family members are listed in annex.

° Special categories of cited documents :

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search

Date of mailing of the international search report

2 November 1999

91.11.99

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Macchia, G

INTERNATIONAL SEARCH REPORT

International Application No

PCT/GB 99/02360

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 98 17687 A (GENETICS INSTITUTE INC (US) JACOBS; MCCOY; LAVALLIE; RACIE; MERBERG ET) 30 April 1998 (1998-04-30) page 6, line 32 -page 8, line 19 page 23, line 12 -page 24, line 14 page 34, line 16-22 page 54, line 17 -page 55, line 5 99% identity between Seq.ID:8 nt.1-1714 and Seq.ID:1 of current application nt.1535-3246 99% identity between Seq.ID:9 and Seq.ID:2 of current application aa.546-1192 page 73 -page 77 page 101 -page 102; claims 20-22 ---	1-12
X	Nucleotide Sequence Database EMBL Entry HSZZ12477; Accession Number AA307342 18 April 1997 100% identity with Seq.ID:1 nt.3013-3395 XP002120596 the whole document ---	1-9,11, 12
P,X	MORRIS N.J. ET AL.: "Cloning and characterization of a 22 kDa protein from rat adipocytes: a new member of the reticulon family" BIOCHIMICA ET BIOPHYSICA ACTA, vol. 1450, no. 1, 6 May 1999 (1999-05-06), pages 68-76, XP002121224 the whole document ---	1-9,11, 12
P,X	WO 98 45435 A (GENETICS INSTITUTE INC (US) JACOBS; MCCOY; LAVALLIE; RACIE; MERBERG ET) 15 October 1998 (1998-10-15) page 56, line 28-34 page 64, line 10-14 page 77, line 15 -page 78, line 2 page 83, line 16-30 100% identity between Seq.ID:1008 and Seq.ID:1 of current application nt.1540-1971 page 428 ---	1-12
P,X	WO 98 45437 A (GENETICS INSTITUTE INC (US) JACOBS; MCCOY; LAVALLIE; RACIE; MERBERG ET) 15 October 1998 (1998-10-15) page 56, line 28-34 page 64, line 10-14 page 77, line 15 -page 78, line 2 page 83, line 16-30 99% identity between Seq.ID:366 and Seq.ID:1 of current application nt.862-1184 page 213 ---	1-12

INTERNATIONAL SEARCH REPORT

International Application No
PCT/GB 99/02360

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	<p>WO 98 56804 A (HUMAN GENOME SCIENCES INC (US) MOORE; SHI; ROSEN; RUBEN; LAFLEUR ET AL) 17 December 1998 (1998-12-17)</p> <p>abstract page 62, line 1-32 Gene No.69 page 82 page 85, line 29 -page 87, line 11 page 95, line 34 -page 96, line 32 page 111, line 6 -page 112, line 16 99% identity between Seq.ID:79 nt.248-814 and Seq.ID:1 of current application nt.3013-3579 page 235 -page 236 100% identity between Seq.ID:179 and Seq.ID:2 of current application aa.1019-1192 page 303 99% identity between Seq.ID:301 and Seq.ID:2 of current application aa.1002-1192 page 354 -page 355 page 370 -page 373; claims ---</p>	1-12
P,X	<p>WO 99 06548 A (GENSET (FR); DUMAS MILNE EDWARDS J.-B.; DUCLERT A.; LACROIX B.) 11 February 1999 (1999-02-11)</p> <p>page 6, line 23 -page 7, line 27 page 78, line 6 -page 79, line 2; example 39 page 84, line 1 -page 85, line 24; example 43 99% identity between Seq.ID:137 nt.165-404 and Seq.ID:1 of current application nt.3013-3252 97% identity between Seq.ID:391 and Seq.ID:2 of current application aa.1002-1084 page 144 -page 148; claims ---</p> <p style="text-align: right;">-/--</p>	1-12

INTERNATIONAL SEARCH REPORT

International Application No
PCT/GB 99/02360

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	<p>WO 99 31236 A (GENSET (FR); BOUGUELERET L.; DUCLERT A.; DUMAS MILNE EDWARDS J.-B.) 24 June 1999 (1999-06-24)</p> <p>abstract</p> <p>page 5, line 21-34</p> <p>page 32, line 3-8</p> <p>page 62, line 21 -page 63, line 5; example 38</p> <p>page 64, line 11 -page 67, line 27; examples 39,40</p> <p>99% identity between Seq.ID:51 nt.68-634 and Seq.ID:1 of current application nt.3013-3579</p> <p>98% identity between Seq.ID:152 and Seq.ID:2 of current application aa.1002-1192</p> <p>page 137 -page 138; claims</p> <p>---</p>	1-12
P,X	<p>WO 99 33979 A (CHIRON CORPORATION (US); LIN H.; CAO L.) 8 July 1999 (1999-07-08)</p> <p>99% identity between Seq.ID:5 nt.687-1253 and Seq.ID:1 of current application nt.3013-3579</p> <p>100% identity between Seq.ID:6 aa.1-185 and Seq.ID:2 of current application aa.1-185</p> <p>page 23, line 5-15</p> <p>page 33, line 8-15</p> <p>page 44, line 4-9</p> <p>page 49, line 24 -page 50, line 10</p> <p>page 59 -page 67; claims</p> <p>---</p>	1-12
P,X	<p>NAGASE T. ET AL.: "Prediction of the coding sequences of unidentified human genes. XII. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro"</p> <p>DNA RESEARCH, vol. 5, no. 5, December 1998 (1998-12), pages 355-364 64, XP002111974</p> <p>ISSN: 1340-2838</p> <p>Gene number KIAA0886</p> <p>the whole document</p> <p>-----</p>	1-5,11, 12

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/GB 99/02360

Patent document cited in search report	Publication date	Patent family member(s)			Publication date
WO 9806841 A	19-02-1998	US 5858708 A			14-01-1999
		AU 3902297 A			06-03-1998
		EP 0918856 A			02-06-1999
WO 9817687 A	30-04-1998	AU 5004097 A			15-05-1998
WO 9845435 A	15-10-1998	AU 6956698 A			30-10-1998
WO 9845437 A	15-10-1998	AU 6956798 A			30-10-1998
WO 9856804 A	17-12-1998	AU 8066798 A			30-12-1998
		AU 6241698 A			07-08-1998
		WO 9831818 A			23-07-1998
		AU 8474398 A			08-02-1999
		WO 9902546 A			21-01-1999
WO 9906548 A	11-02-1999	AU 8554798 A			22-02-1999
WO 9931236 A	24-06-1999	AU 1049199 A			07-06-1999
		AU 1503099 A			05-07-1999
		WO 9925825 A			27-05-1999
		WO 9940189 A			12-08-1999
WO 9933979 A	08-07-1999	AU 1929599 A			19-07-1999